

ABSTRACT

Session identifiers are automatically identified in uniform resource locators (URLs). The session identifiers may be identified using classification techniques based on whether identical sub-strings are identified in multiple URLs downloaded from a web site. The URLs may then have the session identifiers extracted to generate clean versions of the URLs.